DOCUMENT RESUME

ED 238 901 TM 830 841

AUTHOR Kleban, Morton H.
TITLE A Critique of the Standard Method of Interpreting Factor Analysis.
PUB DATE Nov 83
NOTE 11p.; Paper presented at the Scientific Meetings of the Gerontological Society (San Francisco, CA, November 18-22, 1983).
PUB TYPE Speeches/Conference Papers (150) -- Viewpoints (120)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Correlation; *Factor Analysis; *Factor Structure; Multidimensional Scaling; *Regression (Statistics); Research Problems; *Statistical Studies; Test Interpretation
IDENTIFIERS *Stepwise Regression

ABSTRACT

The paper is a critique of the traditional mode of interpreting factor analyses; it is not a criticism of factor analysis per se. Instead, the author proposes a statistical procedure based upon stepwise regression (SRP). The traditional mode focuses on the largest factor loadings (FL). A factor is both described and named by these heavily weighted variables. This method is based on a covert assumption that a factor has a sufficient degree of internal-consistency which is housed in the highly loaded variables. The assumption needs to be tested in every factor. By its very nature, the SRP tests that assumption. Within the SRP, correlations of factor scores with the variables produce the FL's (step 0). At the other end, the multiple regression of all the variables produces a column of unique variance factor-coefficients. The intervening SRP steps provide descriptions of the factors. FL's can enter into many types of configurations. Some examples of the complexity are presented in the paper. A factor is usually composed of several sources of unique and common variances. The traditional method irons out this complexity into a fictitious "oneness." The SRP presents the factor in terms of its independent configurations. (Author)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED238901

A Critique of the Standard Method of Interpreting Factor Analysis

Morton H. Kleban, Ph.D.

Philadelphia Geriatric Center
Philadelphia, PA 19141

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

This document has been reproduced as received from the person or organization originating it.

Minor changes have been made to improve reproduction quality.

Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

M.H. Kleban

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Paper presented at the 1983 Scientific Meetings of the Gerontological Society in San Francisco, CA.

TM 830 841

This paper is a critique of the traditional mode of interpreting factor analysis output; it is not a criticism of the mathematical algorithms but rather of the standard method of interpretation. The paper also proposes another scheme for interpreting factor analyses, one which is based upon the statistical procedure of stepwise regression.

Traditional Mode of Interpretation:

The traditional mode focuses on the largest factor loadings (FL) in the factor structure matrix. The factor is then both described and named with respect to these highly loaded variables. I have interpreted many factor analyses by this method. I never questioned the procedure, but now I am acutely aware of its limitations which are based on a covert assumption of internal-consistency which interpreters rountinely assign to the factors. There is, however, an advantage to the traditional method: it is simple and uncomplicated to apply. On the other hand, the recommendations I shall make for interpretation are far more troublesome and require extra work; but the reward for the additional pain is a far better understanding of the interrelations among variables within factors.

The Complicated Composition of Factors:

I will be focusing on the relationships between the FL and factor coefficient (FC) matrices. The FL's are correlations between the factor and the individual variables; the FC's are partial regression coefficients between the factor (i.e., represented by factor scores) and the variables. In the traditional method, FC's have been used mostly to generate factor scores, but I am suggesting that the FC's be studied more carefully because they identify unique sources of variance within the factor. An array of FL's is just one side of a factor. The other profile is represented by the FC matrix.

When considered from the standpoint of subjects, the term, "unique variance" implies a subset of people (i.e., cases) who are responsible for a relationship. When unique variance is observed in a variable, it means that its significance is associated with a particular subset of subjects. When there are several significant FC's, it means that there are an equal number of independent subsets of people acting within the factor.

With respect to interpretation, however, our major interest is to identify common variance; that is, we want to identify those variables which are linked to a single subset of subjects. These variables are then said to be uni-dimensionally related. The traditional method of interpretation directly addresses itself to that aim. The procedure attempts to insure such common variance by restricting the interpreter's attention only to the highly loaded variables. A substantial FL value, such as .4 or higher, is used to insure that the factor contains a reasonable degree of internal-consistency.

My criticism of the traditional method focuses pointedly on the issue of internal-consistency. Are the highly loaded variables derived from a single source of common variance? In most cases, they are not singularly organized! I will try to demonstrate shortly, through the use of examples, that factor loadings usually derive from multiple sources of both unique and common variances. Of course, the possibility does exist that in a factor all highly loaded variables can be integrated within a uni-dimensional frame. But in my experience, such organizations are exceptions rather than the rule.

The users of the traditonal method, moreover, are usually acutely aware of the importance of the issue of internal-consistency. They often report measurements of internal-consistency (e.g., Cronbach Alpha) to

demonstrate that their factors have a sufficient degree of uni-dimensionality. Such a procedure, however, is highly fallible, and offers no confirmation whatsoever that the factor has a respectable degree of internal-consistency. Such tests require that the variables have a uni-dimensional basis; it is an assumption of the test. The Cronbach Alpha, for example, is sensitive only to the correlative relationships among the variables; it cannot differentiate among uni-dimensional or multi-dimensional, additive sources of influence within the factor.

Stepwise Regression As A Method of Observing Relationships in Factors:

If factor scores are stepwise regressed on the variables (I will be describing the BMDP program as the model), the relationships of the variables to each other and to the factor can be clearly described. In step 0, before starting the regression, one would find the factor loadings re-calculated in the column for partial correlations between the factor and the variables. At step 0, these are zero-order correlations or factor loadings. After all the variables entered into the regression, one would find factor coefficients re-constructed in the B-coefficient column. Thus the origins of the FL's and the FC's become manifest.

The regression steps are filled with meaningful, interpretive information. I shall illustrate the procedure by using examples. First, let us start with an ideal factor that contains uni-dimensional variance; that is, all the variables with high loadings would be produced by a single subset of subjects. Table 1 contains 8 highly loaded variables from a factor analysis of many more variables. The factor loadings are listed in the step 0 column. After step 1, once variable 1, E1, entered the regression, all the factor loadings of the other unentered variables become non-significant statistically. In

this case, the FC's would not have any statistically significant variables. We would be looking exclusively at a circumstance of shared variance. With respect to the issue of uni-dimensionality, however, there may still be a fly in the ointment, but we will examine this difficulty when we analyze table 3.

A multi-dimensional configuration is presented in Table 2. The same factor loadings are listed in step 0. In this example, it would require more regression steps to exhaust the FL variances. At a glance, one can observe a complicated, multi-dimensional factor. And, upon examination, this can be seen to be a very poor factor from the standpoint of common variances. There are only weak sources of such common variance present, and by contrast, there are very powerful sources of unique variance dominating the factor. The factor is essentially a composite of diverse, independent sources of variance. The traditional method of interpretation fails badly on this type of factor. This is not an uncommon configuration, unfortunately. In the traditional method, the FL's appear to be promising; the stepwise regression procedure, however, would easily destroy the illusion of a strongly integrated factor.

Table 3 contains an example which seems most representative of my experience with factors. The factor loadings divide into two distinct sets of shared variance. The first source of common variance, associated with regression step 1, can be found in the variables 1, 3, 4, 5, and 6. With the entrance of Es, a second, independent source of common variance is revealed, focused in the variables, 2, 7, and 8. In this example, we observe two separate sources of common variance. The traditional method would blindly link them together. It is obvious, that they are formed by two different subsets of subjects.

Now I must return to the fly in the ointment. The stepwise regression can be used as a powerful aid to helping us observe the organizations within a factor. But the procedure does not necessarily identify uni-dimensional sources of common variance. If an entering variable has two distinct influences, the interpreter would not be able to differentiate these influences in the associated, unentered variables. If the E1 variable in Table 1 contained relationships with two distinct subsets, they would both enter simultaneously. A portion of variance in E1 may pull shared variance from variables 3, 5, and 7. Another part of E1 could pull in the remaining variables. Although there were no statistically significant FC's, nevertheless the factor might have a dichotimous rather than a uni-dimensional base.

It should be quite obvious, that uni-dimensionality is a difficult state to demonstrate. If one is interested in observing such pure clusters, however, then it would be necessary to do further statistical regressions to isolate these relationships. The problem might be solved with an alternate stepwise sequence of the entering variables; or it might be necessary to do specialized statistical analyses to ferret the problem. On many occasions, however, the interpreters might not be interested in such observational refinements.

Although the stepwise regression procedure requires much more work than the traditional method, it takes the analyst closer to the factorial "truths." The interpreters would be able to describe the factors with much greater clarity, and he or she would be able to dispense with the ironing board interpretations of factors which result from the use of the traditional method.

Table 1

Stepwise Regression of a Factor with a Uni-dimensional basis

Variables	Factor Loadings Regression Steps	
	0	1
Variable 1	.67	E
Variable 2	.59	.05
Variable 3	.65	.15
Variable 4	.65	.10
Variable 5	.53	.09
Variable 6	.55	.16
Variable 7	.52	.12
Variable 8	.60	.22

Table 2

Stepwise Regression of a Factor with a Unique variance, Multi-dimensional Basis

Variables	Factor Loadings Regression Steps							
	0	1	2	3	4	5	6	7
Variable 1	.67	E1						
Variable 2	.59	.51	E2					
Variable 3	.65	.50	.50	E3				
Variable 4	.65	.50	.49	.42	E4			
Variable 5	.53	.44	.38	.37	.36	E5	E6	
Variable 6	.55	.43	.38	.34	.29	.28	.33	E7
Variable 7	.52	.42	.43	.35	.23	.25	.27	.30
Variable 8	.60	.43	.36	.32	.28	.27		

Table 3

Stepwise Regression of a Factor with Two Sources of Common Variance

Variables	Factor Loadings Regression Steps		
	0	1	2
Variable 1	.67	E1	
Variable 2	.59	.51	E2
Variable 3	.65	.05	.05
Variable 4	.65	.05	.05
Variable 5	.53	.05	.05
Variable 6	.55	.05	.05
Variable 7	.52	.42	.02
Variable 8	.60	.43	.02

ABSTRACT FORMS: *Must be Postmarked by April 15, 1983*

PLEASE READ AND FOLLOW INSTRUCTIONS CAREFULLY
(Note: Any abstract unsuitable for direct reproduction must be retyped—the charge to the author is $10.00)

Use full space for summaries of Symposia, Workshops, and Discussion Sessions — Use space above dotted line for Papers, Posters, Media and Individual Symposia Presentations

A CRITIQUE OF THE STANDARD METHOD OF INTERPRETING FACTOR ANALYSES. Morton H. Kleban, Philadelphia Geriatric Center, Phila.,PA 19141.

The paper is a critique of the traditional mode of interpreting factor analyses; it is not a criticism of factor analysis per se. Instead, the author proposes a statistical procedure based upon stepwise regression (SRP). The traditional mode focuses on the largest factor loadings (FL). A factor is both described and named by these heavily weighted variables. This method is based on a covert assumption that a factor has a sufficient degree of internal-consistency which is housed in the highly loaded variables. The assumption needs to be tested in every factor. By its very nature, the SRP tests that assumption. Within the SRP, correlations of factor scores with the variables produce the FL's (step 0). At the other end, the multiple regression of all the variables produces a column of unique variance factor-coefficients. The intervening SRP steps provide descriptions of the factors. FL's can enter into many types of configurations. Some examples of the complexity are presented in the paper. A factor is usually composed of several sources of unique and common variances. The traditional method irons out this complexity into a fictitious "oneness." The SRP presents the factor in terms of its independent configurations.

APPLICATION FOR SECTION SPONSORSHIP

(Symposia, discussions and workshops may be sponsored by more than one section: for papers or posters CHECK ONLY ONE BOX.)

- ☐ Biological Sciences (000)
  (☐ Submitted for BS Section student award)
- ☐ Clinical Medicine (001)
- ☐ Social Research, Planning and Practice (005)

Behavioral and Social Sciences
- ☒ *Psychology* (002)
- ☐ *Sociology* (003)
- ☐ *Other Social Sciences* (004)

MEMBERSHIP STATUS

Status of first author or organizer; check one.

- ☒ Member (006)
- ☐ Nonmember (008)
- ☐ Student Member (007)
- ☐ Student Nonmember (009)

SUBJECT AREA OR FIELD OF INTEREST

Please enter in the space below the appropriate code number of the subject area or field of interest which you feel most accurately describes your abstract. See pp. 9-10.

Code No. 089

PRESENTATION PREFERENCE

Preferred Choice (Check one)

- ☒ Paper Session (300)
- ☐ Poster (301)
- ☐ Media (306)
- ☐ Symposium (302)
- ☐ Workshop (303)
- ☐ Discussion (304)
- ☐ Round Table (305)

If first choice is unavailable, I will:

- ☒ Accept an alternative
- ☐ Withdraw the abstract

AUDIO-VISUAL EQUIPMENT

Regular Sessions & Media Room

All equipment must be ordered in advance. *Only* equipment specified here will be available at the time of the meeting.

Please check one:

- ☐ 2×2 sl. proj. (380)
- ☐ 16 mm sound proj. (381)
- ☐ 8 mm movie proj. (382)
- ☐ Electric Pointer (383)
- ☐ Overhead (384)
- ☐ Audio tape recorder (385)
- ☐ No equipment (386)
- ☐ Videotape—will be provided for the special Media Room only. Not provided for regular sessions—see page 2. (387)

REVIEWERS' COMMENTS: